

TR-0201-02

**Neurophysiological Based Methods of  
Guided Image Search**

**Final Report**

10 February 2003

Prepared for:

National Imagery and Mapping Agency  
4600 Sangamore Road  
Bethesda, MD 20816-5003

Under Contract:

NMA401-02-C-0007

Prepared by:

Frank M. Marchak, Ph.D.

20030616 067

**Veridical Research and Design**

PO Box 6503  
Bozeman, Montana 59771-6503

| REPORT DOCUMENTATION PAGE  |   |  | Form Approved<br>OMB No. 0704-0188                        |
|--|---|--|---|
| <p>Pulic reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</p>  |   |  |   |
| 1. AGENCY USE ONLY (Leave Blank)   | 2. REPORT DATE<br>2/11/2003                                 | 3. REPORT TYPE AND DATES COVERED<br>Final Report 10Apr02-11Feb03 |   |
| 4. TITLE AND SUBTITLE<br>Neurophysiological Based Methods of Guided Image Search: Final Report   |   |  | 5. FUNDING NUMBERS<br>C-NMA401-02-C-007                   |
| 6. AUTHOR(S)<br>Frank M. Marchak, Ph.D.  |   |  |   |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>Veridical Research and Design<br>PO Box 6503<br>Bozeman, MT 59771-6503   |   |  | 8. PERFORMING ORGANIZATION<br>REPORT NUMBER<br>TR-0201-02 |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br>National Imagery and Mapping Agency<br>12310 Sunrise Valley Dr.<br>Reston, VA 20191   |   |  | 10. SPONSORING/MONITORING<br>AGENCY REPORT NUMBER         |
| 11. SUPPLEMENTARY NOTES  |   |  |   |
| 12a. DISTRIBUTION/AVAILABILITY STATEMENT<br>Approved for public release; distribution unlimited.   |   |  | 12b. DISTRIBUTION CODE                                    |
| 13. ABSTRACT (Maximum 200 words)<br><p>Report developed under SBIR contract for topic NIMA02-001. Complex analysis of intelligence imagery is crucial to the missions of intelligence organizations, yet remains constrained by labor-intensive, time-consuming visual search of large volumes of imagery. Many algorithms have been developed to automatically identify regions of interest in large, complex sets of imagery, yet the utility of such algorithms is limited by the fact that human analysts detect features in imagery with higher accuracy than existing methods. We developed a model of visual feature detection, the Neuronal Synchrony Model, based on neurophysiological models of temporal neuronal processing, to improve the accuracy of automatic detection of features of interest in complex natural imagery. The Neuronal Synchrony Model of image feature detection was applied to accurately identify and highlight regions of images that contain target features, thus automating the labor-intensive, "scanning" portion of imagery analysis. The accuracy of the Neuronal Synchrony Model was tested with natural images containing visually controlled, synthetic targets as well as with natural targets using a variety of overhead imagery background and target types. A proof-of-concept demonstration of the effectiveness of this model showed enhancement in the speed and accuracy of interactive, guided visual search of particular classes of imagery.</p> |   |  |   |
| 14. SUBJECT TERMS<br>SBIR Report, Neuroscience modeling, imagery analysis, psychophysics, visual search  |   |  | 15. NUMBER OF PAGES                                       |
|  |   |  | 16. PRICE CODE  |
| 17. SECURITY CLASSIFICATION OF<br>REPORT<br>Unclassified   | 18. SECURITY CLASSIFICATION OF<br>THIS PAGE<br>Unclassified | 19. SECURITY CLASSIFICATION OF<br>ABSTRACT<br>Unclassified       | 20. LIMITATION OF ABSTRACT<br>UL                          |



## Table of Contents

|   |           |
|---|-----------|
| <b>INTRODUCTION .....</b>                                 | <b>3</b>  |
| <b>BACKGROUND .....</b>                                   | <b>3</b>  |
| <b>TECHNICAL APPROACH .....</b>                           | <b>5</b>  |
| <b>OBJECTIVES.....</b>                                    | <b>5</b>  |
| <b>TASK 1 – NEURONAL SYNCHRONY MODEL DEVELOPMENT.....</b> | <b>5</b>  |
| <b>MODEL DESIGN.....</b>                                  | <b>5</b>  |
| MODEL CELLS.....  | 6         |
| NETWORK FILTERS .....                                     | 6         |
| TEMPORAL SYNCHRONIZATION .....                            | 9         |
| <b>MODEL IMPLEMENTATION.....</b>                          | <b>10</b> |
| MODEL CELLS.....  | 10        |
| NETWORK .....   | 11        |
| TEMPORAL SYNCHRONIZATION .....                            | 11        |
| <b>TASK 2 – NEURONAL SYNCHRONY MODEL EVALUATION .....</b> | <b>12</b> |
| <b>FUNCTIONAL EVALUATION .....</b>                        | <b>12</b> |
| <b>HUMAN VISUAL SEARCH PERFORMANCE EVALUATION.....</b>    | <b>12</b> |
| TEST IMAGE CREATION.....                                  | 12        |
| PSYCHOPHYSICAL TESTING – SYNTHETIC IMAGERY.....           | 15        |
| <b>TASK 3 – MODEL REFINEMENT.....</b>                     | <b>22</b> |
| <b>MODEL APPLICATION TO EXTENDED IMAGERY SET .....</b>    | <b>22</b> |
| <b>UPDATE AND REFINEMENT .....</b>                        | <b>24</b> |
| <b>MODEL DESIGN.....</b>                                  | <b>24</b> |
| <b>MODEL IMPLEMENTATION.....</b>                          | <b>25</b> |
| PSYCHOPHYSICAL TESTING – OVERHEAD IMAGERY .....           | 27        |
| <b>SUGGESTIONS FOR FUTURE RESEARCH.....</b>               | <b>32</b> |
| <b>REFERENCES .....</b>                                   | <b>33</b> |



## INTRODUCTION

To support NIMA's goals of improving analysts' productivity and automating tasks currently performed by analysts, Veridical Research and Design was engaged in an effort to determine the feasibility of developing a neurophysiologically based methodology that identifies and highlights potential contours of interest in overhead imagery to guide the image analyst (IA) in visual exploitation tasks. The objective of this effort was to research and develop a concept, based on established neurophysiological and psychophysical knowledge of how the human visual system segments scenes and detects perceptual groupings of features, that will allow analysts to complete search tasks in a shorter period of time and with greater confidence in their analysis.

Accurate and timely visual analysis of imagery is of concern to government, industry and academic researchers across a variety of domains and applications. Search tasks are crucial to the missions of many exploitation organizations, yet are among the most difficult. The volume of imagery, the relatively low degree of intelligence content and the competing demands on the analyst's time all contribute to a less than optimal search process. The basic problem in image exploitation is that the reduced information contained in still, monochrome images allows little or no fast discrimination between objects of interest and everything else in the image. As a result, image exploitation is inefficient as the IA is forced to scrutinize everything. Under these circumstances, increases in efficiency can be achieved only with decreased report confidence. The concept described here is a methodology to identify areas of an image that are likely to contain contours of interest, thereby quickly guiding the IA's attention to high probability target areas so that the total region that must be scrutinized is substantially reduced. The result is improved allocation of attentional resources that increases analysis confidence and accuracy.

While a variety of mathematically derived image processing algorithms have been developed to accomplish this task, to date few methods have been found that exceed the visual abilities of human analysts in the assessment and evaluation of imagery, particularly of natural scenes. The novel approach to imagery analysis described here is based on the Neuronal Synchrony Model and supported by neurophysiological research of contour perception in the mammalian visual system. Specifically, it involves differentiating perceptually relevant contours from a noisy background using a Neuronal Synchrony Model of the striate cortex. The resultant potential contours of interest are then highlighted in a manner that augments the analyst's own visual processing of the image, thus decreasing workload and increasing the precision of image exploitation tasks. The output of this effort is a proof-of-concept demonstration of the effectiveness of this methodology in enhancing visual searches on representative imagery. The next subsection provides background information on current methods of imagery analysis and outlines the neurophysiological research that inspires the Neuronal Synchrony approach to imagery analysis.

## BACKGROUND

Image exploitation involves a variety of techniques and methods. A typical processing strategy often entails a full scene 'reconnaissance' processing to locate particular regions of interest and then an in-depth analysis of the selected region to detect and identify specific objects of interest (Guindon, 1997). One approach to automating this process requires seeking local salient contours in an image and then seeking target objects in the image based on detected contour groupings. One of the most difficult problems, regarded as a bottleneck in this bottom-up process, is caused by the fact that low-level feature extraction algorithms often produce less than optimal results. Application of traditional computer vision methods such as edge detection to extraction of information from digital remotely sensed images is complicated by the characteristics of aerial and satellite imagery, including noise resulting from the acquisition process, variations in illumination and geometry due to camera angles, as well as the effects of shading (Sowmya &





Trinder, 2000). Similarly, conventional neural network approaches also tend not to perform well on such imagery because of the high degree of partial and noisy information (Raghu & Yegnanarayana, 1997). As mentioned previously, human analysts excel at the ability to accurately detect features of interest in imagery. None of the current mathematical edge detectors have psychophysical correlates in a manner, for example, that the CIE  $L^*a^*b^*$  color model corresponds with human color perception (Wyszecki & Stiles, 1982). Examination of how the human visual system accomplishes these tasks at the perceptual and neurophysiological levels can provide insights to approaches that can be applied to better addressing this issue.

Physiological investigations in human and other mammalian visual systems have shown that simple cells in the visual cortex of the brain are sensitive to intensity gradients of particular orientations and spatial frequencies across the visual field (Hubel & Wiesel, 1959). Thus, one of the earliest cortical representations of a visual scene is based on multi-resolution and multi-orientational receptive fields. Further, it has also been shown that these receptive fields are well described by Gaussian modulated sinusoid, or "Gabor," functions (Marcelja, 1980). A mechanism employing these characteristics could conceivably produce results in visual perception tasks similar to that of human performance.

One source of insights toward such a mechanism comes from anatomical and neurophysiological studies of processing in the striate cortex. Specifically, Gilbert, Das, Ito, Kapadia & Westheimer (1996) have proposed that cortical cells integrate information over large portions of the visual field through long-range horizontal connections and that distant surrounding features can thereby modulate the response of these cells. In addition, it has been shown that these same cortical cells are involved in neuronal temporal synchronization (Singer & Gray, 1995). Temporal synchronization holds that signals of neurons that are to be grouped together are correlated in time. Signals are evaluated by determining the temporal correlations of signal fluctuations occurring over a particular time scale. A model that embodies units with characteristics of cortical cells in a network of long-range connections and capable of temporal synchronization could well capture many of the characteristics of human contour perception (Yen & Finkel, 1998).

Several models have been developed that express visual segmentation in terms of neuronal firing correlations (Sporns, Tononi, & Edelman, 1991; von der Malsburg & Buhmann, 1992; Yen & Finkel, 1998). These models employ permanent excitatory connections between neurons to encode the likelihood of being activated by the same figure, reflecting the Gestalt laws of perceptual grouping. Examples of these connection types include the connection between two neurons in different locations in the visual field that correspond to the same local feature (e.g. intensity, orientation) or excitation between neurons corresponding to the same location in input space. By synchronizing neuronal activity in time, the combined action of all excitatory connections binds neurons together that are activated by the same figure. In a similar manner, inhibitory connections suppress simultaneous activity in cells belonging to figure and ground, making those signal correlations unambiguous and reducing accidental coincidences. The result is that the different excitatory connections create global correlations within a figure or background and provide a basis for feature combinations required for scene segmentation (von der Malsburg, 1999).

There is a growing body of evidence supporting the efficacy of the temporal correlation hypothesis as a model of visual feature integration, both neurophysiologically (Schmidt, Goebel, Lowel, & Singer, 1997; Singer, 1999) and computationally (Deco & Zihl, 2000; Sporns et al., 1991). Unlike traditional neural net approaches that have a fundamental combinatorial coding problem in dealing with the issue of feature binding (von der Malsburg, 1999), models based on temporally correlated cell assemblies circumvent this difficulty and accurately reflect human performance (Hummel & Biederman, 1992). Taken together, these findings suggest that a model employing a mechanism based on temporal correlation might perform at levels similar to that of the human visual system and do a better job at identifying contours of interest than traditional image analysis algorithms. The next section details our approach for employing this model for guided imagery exploitation.



## TECHNICAL APPROACH

### OBJECTIVES

The primary objective of this effort was to develop a neurophysiologically based model that characterizes the perceptually relevant contours of interest in overhead imagery of natural scenes in a manner consistent with human performance, utilizing only information available in the image. The goal was to apply the model to highlight contours of potential interest in a manner that augments the analyst's own visual processing. The output of this work is a proof-of-concept demonstration for enhancing the analyst's ability to perform searches on representative imagery.

The specific technical objectives of Phase I include:

- Neuronal Synchrony Model Development
  - Development of an initial Neuronal Synchrony Model of contour detection and creation of test imagery
- Model Evaluation
  - Evaluation of the performance of the model and determination of its effectiveness in supporting a more effective visual search process
- Model Refinement
  - Application of the model to a larger set of representative test imagery and updating and tuning performance based on the results.

The sections below describe the overall methodology and results of the effort to create a neuronal synchrony model of image analysis. The information below provides an overall summary of the findings as well as providing suggestions for further research. For detailed information regarding results of preliminary efforts on these tasks, see Marchak (2002).

### TASK 1 – NEURONAL SYNCHRONY MODEL DEVELOPMENT

The primary goal of this task was to develop an initial Neuronal Synchrony Model, based on the work of Yen & Finkel (1998). The model determines contours of perceptual relevance in imagery and outputs the results for use in highlighting these regions for subsequent processing by an analyst. A secondary goal of this task was to create test imagery that could be used to evaluate the performance of the model in the next task.

Yen and Finkel (1998) have developed a model – based on temporal synchronization – for computing the perceptual salience of contours embedded in noisy images and have shown the effectiveness of the approach in accounting for a range of basic psychophysical and neurophysiological phenomena. In coordination with Dr. Charles Gray and Dr. Shih-Cheng Yen, we have been extending and expanding their methodology and incorporating it in a system that will use the output of the model to guide analysts to probable contours of interest in remotely sensed imagery.

### MODEL DESIGN

The basic premise of the Neural Synchrony Model is that perceptual salience arises from the temporal synchronization of cells with long-range horizontal connections in the striate cortex. The model developed



here is composed of Gabor-based units modeled on striate cortex cells that are embedded in a network made up of long-range horizontal connections and capable of temporal synchronization (Yen & Finkel, 1998). The implementation of the model involved developing model cells based on Gabor-based units, creation of network filters that implemented facilitation and inhibition of the long-range horizontal connections between units, and development of a temporal synchrony mechanism to determine perceptual salience. Each of these factors is discussed below.

## MODEL CELLS

As outlined in our proposal, Gabor wavelet transforms were chosen to model the responses of cortical cells. The wavelet transform provides a mathematical technique for hierarchically decomposing an image into multiresolution components that provide a description of the coarse, overall texture, as well as details that range from broad to narrow (Mallat, 1999). The resulting set of coefficients can be used to uniquely define an image and determine specific features across a variety of scales and orientations. Of particular interest is the fact that the same characteristics of simple cells in the mammalian primary visual cortex – spatial localization, orientation sensitivity, and band pass preference – are comparable to the basis functions of wavelet transforms (Olshausen & Field, 1996). The wavelets essentially play the role of receptive fields with the response of the neuron captured by the wavelet coefficients.

The response of the Gabor filters used in the model are represented as:

$$A(x, y, \theta) = \iint f(x, y, \theta) I(x, y) dx dy \quad (1)$$

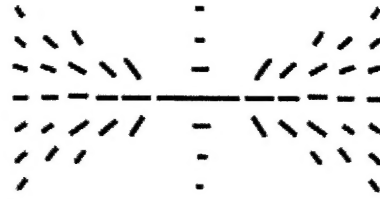
where  $f(x, y, \theta)$  is the kernel of the steerable Gabor filter oriented at  $\theta$  and  $I(x, y)$  is the input image.

## NETWORK FILTERS

As described in Yen and Finkel (1998), the model cells are embedded in a network of long-range horizontal connections that are capable of temporal synchronization. Under the direction of Dr. Yen, we developed a MATLAB implementation of this network using facilitation and inhibition properties based on psychophysical and neurophysiological evidence of human perception ((Field, Hayes, & Hess, 1993; Gray, 1999; Kovacs & Julesz, 1993; Polat & Sagi, 1993).

## FACILITATION

Each model cell receives weighted inputs from other oriented cells in its surround dependent on the position and orientation of the pre-synaptic cells, resulting in separate spatial zones of excitation and inhibition. The excitatory connections are made up of two regions, a co-axial region which spreads out along the orientation axis of the cell and a trans-axial region that extends orthogonal to the cell orientation. Figure 1 from Yen and Finkel (1998) shows the connection pattern of a horizontally oriented cell, with the orientation and length of the lines representing the “preferred” orientation and connection strength, respectively, as outlined below.



**Figure 1. Connectivity Pattern of Horizontally Oriented Cell**

The co-axial connections consist with Field's et al. (1993) "association field" derived from results of human psychophysical performance and are based on the "co-circular" connection scheme of Parent and Zucker (1989). The co-circularity rule states that for two points lying on a circle, the average of their tangent orientations equals the slope of the line between them. For a cell of some orientation  $\theta_A$  at location A, there is a "preferred" orientation at location B,  $\phi_B$ , described by this rule. Local orientation activity distribution at B strongly facilitates  $\theta_A$  when  $\phi_B$  peaks, and decreases as the orientation of B deviates from  $\phi_B$ . Given a post-synaptic cell at orientation  $\theta$ , the preferred orientation  $\phi$  at position  $(i,j)$  of the pre-synaptic cell  $\psi$  is described by:

$$\phi(\theta, i, j) = 2 \tan^{-1} \left( \frac{j}{i} \right) - \theta \quad (2)$$

The connection weights peak at  $\phi$ , falling off as a Gaussian function of the difference in the acute angle ( $| \text{acute} |$ ) between  $\phi$  and  $\psi$ , with the half-width at half-height  $\sigma_\psi^c$  for the co-axial connections specified by:

$$B(\theta, i, j, \psi) = G(|\psi - \phi(\theta, i, j)|_{\text{acute}}, \sigma_\psi^c) \quad (3)$$

In this scheme, connection weights decrease with increasing angular deviation from the cell orientation, which further reflects the results of human visual performance that show a preference for lines that are straight and of low curvature (Field et al., 1993). In addition, in accordance with the psychophysical results of Polat and Sagi (1993), connection weights decrease as a function of distance:

$$D(i, j) = G(\sqrt{i^2 + j^2}, \sigma_d^c) \quad (4)$$

where  $\sigma_d^c$  is the half-width at half-height of the Gaussian function.

The trans-axial excitatory connections extend orthogonally from the cell with a more spatially focused pattern, with the strongest connections closest to the cell. These connections are calculated in manner



similar to the co-axial connections using  $\sigma_\psi^t$  and  $\sigma_d^t$ . The co-axial and trans-axial connections compete so that only one set is active at a time, permitting the stimuli to modulate the dominance of the connections.

As mentioned above, the fan-out of the connections is limited to low curvature deviations from the orientation axis as well as an orthogonal region. A further constraint is that the horizontal connections are reciprocal, based on neurophysiological evidence (Kisvarday & Eysel, 1992). These factors are addressed in model as:

$$\Gamma(\theta, i, j, \psi) = \begin{cases} 1, \text{if } \tan^{-1}\left(\frac{j}{i}\right) - \theta < \kappa^c, \\ \text{and } \tan^{-1}\left(\frac{j}{i}\right) - \psi < \kappa^c \\ 1, \text{if } \tan^{-1}\left(\frac{j}{i}\right) - \theta = \frac{\pi}{2} \pm \kappa^t, \\ \text{and } \tan^{-1}\left(\frac{j}{i}\right) - \psi = \frac{\pi}{2} \pm \kappa^t \\ 0, \text{otherwise} \end{cases} \quad (5)$$

where  $\kappa^c$  is the maximum angular deviation of the co-axial connections and  $\kappa^t$  is the maximum angular deviation of the trans-axial connections. Reciprocal fan-out constraints imposed by the pre-synaptic cell are represented by  $(-i, -j)$ .

Taken together, these representations describe the facilitation for a cell at orientation  $\theta$  located at position  $(x, y)$  as:

$$F(x, y, \theta) = \iiint_{(i, j) \in N} \Gamma(\theta, i, j, \psi) D(i, j) B(\theta, i, j, \psi) A(x+i, y+j, \psi) d\psi di dj \quad (6)$$

## INHIBITION

The primary function of inhibition is to differentiate signal from noise based on the level of facilitation. Random elements in the background may by chance be optimally oriented with a contour element. Inhibition serves to prevent these elements from attaching to a target contour. The facilitation stage of the model extracts cells receiving input above the level of noise. This stage implements a longer-latency inhibition that arises from outside the facilitatory zones:

$$I(x, y, \theta) = \begin{cases} 1, \text{if } \tan^{-1}\left(\frac{i}{j}\right) - \theta > \kappa^c, \\ 1, \text{if } \tan^{-1}\left(\frac{j}{i}\right) - \theta > \frac{\pi}{2} \pm \kappa^t, \\ 0, \text{otherwise} \end{cases} \quad (7)$$



The magnitude of the inhibition suppresses any cell with total support less than a fixed threshold so that it is suppresses only cells with weak support but not affecting strongly facilitated cells.

## TEMPORAL SYNCHRONIZATION

As described above, neurophysiological evidence suggests that the same cortical cells interconnected by long-range horizontal connections are also involved in temporal synchronization (Gray, 1999). In the model, the chattering cells are represented as homogenous-coupled neural oscillators, with the phase of each oscillator modulated by the phase of other coupled oscillators (Terman & Wang, 1995). Each oscillator encodes some feature of an object, even just a pixel. Each object segment is represented by a group of oscillators that show synchrony – phase-locking with zero phase shift – with different groups of oscillators desynchronized from each other representing different objects.

Initial efforts involved use of the neural oscillator model described in (Terman & Wang, 1995; Wang & Terman, 1995). The model is based on locally excitatory, globally inhibitory oscillator networks (LEGION). A single oscillator,  $i$ , is defined as a feedback loop between an excitatory unit  $x_i$  and an inhibitory unit  $y_i$ :

$$\frac{dx_i}{dt} = 3x_i - x_i^3 + 2 - y_i + I_i + S_i + \rho \quad (8a)$$

$$\frac{dy_i}{dt} = \varepsilon [\gamma(1 + \tanh(x_i / \beta)) - y_i] \quad (8b)$$

where  $I_i$  denotes external stimulation to the oscillator,  $S_i$  represents coupling from other oscillators in the network, and  $\rho$  represents the amplitude of the Gaussian noise term. Noise is introduced to actively desynchronize different input patterns.

As shown in Figure 2, reproduced from Wang and Terman (1995), the  $x$ -nullcline of (8) is a cubic curve and the  $y$ -nullcline is a sigmoid curve. When  $D > 0$ , the curves intersect along the middle branch of the cubic and (8) is oscillatory, with the periodic solution alternating between silent and active phases of near steady-state behavior.

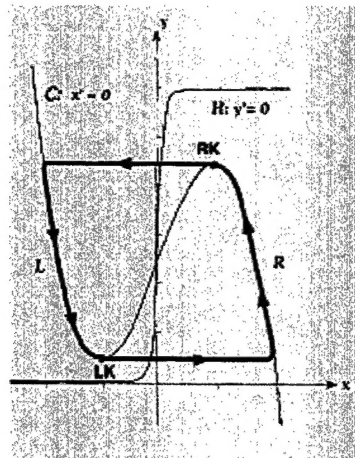


Figure 2. Nullclines and periodic orbit of single oscillator.  
Reproduced from Wang and Terman (1995).





The parameter  $\gamma$  in (8b) controls the relative times that the solution alternates between silent and active phases of steady-state behavior, while  $\beta$  controls the steepness of the sigmoid. The coupling term  $S_i$  is described by:

$$S_i = \sum_{k \in N(i)} W_{ik} S_\infty(x_k, \theta_x) - W_z S_\infty(z, \theta_z) \quad (9)$$

where

$$S_\infty(x, \theta) \equiv \frac{1}{1 + \exp\{-\kappa(x - \theta)\}} \quad (10)$$

$W_{ik}$  describes the connection weight from oscillator  $k$  to oscillator  $i$ , and  $N(i)$  represents the set of adjacent oscillators that connect to  $i$ .  $\theta_x$  controls the threshold for which an oscillator can affect its neighbors.  $W_z$  represents the weight of inhibition from the global inhibitor  $z$ , with activity represented by

$$\frac{dz}{dt} = \phi(\sigma_\infty - z) \quad (11)$$

where  $\sigma_\infty = 0$  when  $x_i < \theta_x$  for every oscillator, and  $\sigma_\infty = 1$  if  $x_i \geq \theta_x$  for at least one oscillator  $i$ . Therefore, once an oscillator is in the active phase, it triggers the global inhibitor, which inhibits the entire network as shown in (9). However, an active oscillator spreads activation to nearest neighbors and onto further neighbors.

## MODEL IMPLEMENTATION

### MODEL CELLS

A Gabor-based algorithm developed by Nestares, Navarro, Portilla & Taberner (1998) was chosen to represent the model cells. This approach employs a multiresolution pyramid using Gabor basis functions. These Gabor wavelets are complex exponential functions modulated by Gaussian functions. These wavelets can be tuned to a continuum of spatial positions, frequencies and orientations. From the basic Gabor functions, the complete set of functions used for sampling the joint space-frequency domain is obtained using three manipulations. Rotations, with a  $45^\circ$  step, are used to get four orientation channels in the frequency domain. Stretching by a factor of two to halve the frequency of the sinusoid permits sampling the frequency domain in octaves. Translations in horizontal and vertical locations are used to sample the spatial domain.

A MATLAB realization based on this algorithm was obtained and implemented in our computing environment. Using four spatial frequencies, the coefficients of the Gabor transform were obtained through convolution of an image with each one of the filters. The distribution by octaves of the filters' tuning frequencies was implemented as a multiresolution pyramid. The same set of filters was applied to low-pass filtered and down sampled versions of an original image. This pyramidal implementation produces a redundant but robust representation that permits reconstruction by adding together all of the even Gabor channels plus their residuals. The resulting algorithm permits calculation of Gabor wavelet coefficients at four different scales and four different orientations that serve as the cortical units in the Neuronal Synchrony Model. In addition, the reconstruction functions supported test image creation as described in the Model Evaluation section. Complete details of the Gabor decomposition procedures can be found in (Marchak, 2002).





## NETWORK

As described in Yen and Finkel (1998), the model cells are embedded in a network of long-range horizontal connections that are capable of temporal synchronization. Under the direction of Dr. Yen, we developed MATLAB implementations of this network using facilitation and inhibition properties based on (6) and (7).

Facilitation filters were developed for each of the four orientations (0, 45, 90, 135 degrees) using long-range horizontal connections weighted by both the preferred angle and distance from center. The preferred angle,  $\phi$ , was calculated using (2). This value is a function of the difference between  $\theta$ , the orientation of the filter - which represents the orientation of the post-synaptic cell - and  $\psi$ , the orientation of the input image - which represents the pre-synaptic cell. Next connection weights,  $B$ , which peak at  $\phi$  and fall off as a Gaussian function of the difference in the acute angle between the  $\phi$  and  $\psi$ , were calculated using (3). For the co-axial connections, distance,  $\sigma_d^c$ , was set to 40, while the angle of the co-axial connections,  $\sigma_\psi^c$ , was 20 degrees. For the trans-axial connections,  $\sigma_d^t$ , the distance of the trans-axial connections, was 30 and  $\sigma_\psi^t$ , the angle of the trans-axial connections, was 20 degrees. In a similar manner, distance weights,  $D$ , were calculated using (4). Lastly, fan-out of the connections,  $\Gamma$ , which limits connections to low curvature deviations from the orientation axis and to a narrow region orthogonal to the orientation axis, was determined by applying (5). The fan-out of the co-axial connections,  $\kappa^c$ , was 30 degrees and the fan-out of the trans-axial connections,  $\kappa^t$ , was 10 degrees.

Facilitation was then determined by combining  $B$ ,  $D$ , and  $\Gamma$ . In this manner, a facilitation filter was calculated for an input image and filter orientations of 0, 45, 90, and 135 degrees. These facilitation filters were applied to the associated Gabor output for each image. The resulting output contained a representation of cell activity at each pixel for a particular orientation based on facilitation of cells in its surround.

Inhibition, based on (6), was then applied to the output of facilitation to suppress cells with weak facilitation. Inhibition was implemented by assuming that each element divides its support equally among cells that provide suprathreshold pre-synaptic input. Any cell whose input was less than a fixed threshold, 0.5, was suppressed. The outputs of the facilitation and inhibition processes for each of the four orientations were then recombined into a single image. This image was then submitted to temporal synchronization as described below.

## TEMPORAL SYNCHRONIZATION

A two-dimensional LEGION oscillator network was implemented in MATLAB based on the equations in Wang and Terman (1995) described above and as implemented in a C program developed by Wang (1994). The oscillator network developed here consisted of a 1000 x 1000 grid of oscillators, based on the input size of the imagery used in the evaluation below, with a global inhibitor as defined in (8) - (11), with the differential equations solved using a fourth-order Runge-Kutta method. Each oscillator in the simulation was connected to its nearest neighbor by a constant weight set to 2.5 for simplicity. The remaining parameters were set based on the values used by Wang and Terman (1995):  $\epsilon=0.02$ ,  $\phi=3.0$ ,  $\gamma=6.0$ ,  $\beta=0.1$ ,  $\kappa=50$ ,  $\theta=-0.5$ ,  $\theta_{xx}=\theta_{xy}=0.1$ ,  $\rho=0.02$ .

The LEGION implementation begins by calculating the connection weights for the eight immediate neighbors of each oscillator. Next, the initial phase of each oscillator is randomly initialized. The net input for each oscillator is calculated over all inputs, the strongest link is traversed, and oscillators that jump into synchronization are found. All pixels in a region defined by an oscillator are then assigned the same numerical value, beginning at 1 and continuing on through the number of regions, up to 255.



The output of the LEGION algorithm is a pixel map that assigns a unique number to each region defined in the input region defined by the correlated oscillations. The resulting region values were each mapped to saturation levels of a yellow hue to produce highlighted regions. All other pixels were assigned to black. The highlights from the oscillator output were then overlaid on the original imagery to indicate perceptually salient regions. The next section provides examples of this process and discusses model evaluation.

## **TASK 2 – NEURONAL SYNCHRONY MODEL EVALUATION**

The primary goal of this task was to evaluate the model's performance, both in terms of basic functionality and overall effectiveness. This was accomplished through functional evaluation of the individual model components, as well as by evaluating model performance in relation to its ability to facilitate visual search in imagery.

### **FUNCTIONAL EVALUATION**

Each of the separate components of the Neuronal Synchrony Model was evaluated to determine that the output accurately matched the design. The individual elements of the facilitation filter – B, D, and  $\Gamma$  – and the inhibition component were compared with manual calculations of a subset of the inputs. Any discrepancies were investigated and corrected. The LEGION model implementation was similarly evaluated, first by comparing output to artificial stimuli containing separate, distinct regions and then by application to the imagery to be used for human performance testing. The results were accurately functioning modules that carried out the calculations necessary for the human visual search performance testing.

### **HUMAN VISUAL SEARCH PERFORMANCE EVALUATION**

In order to evaluate the effects of the model on human visual search performance, a set of visual stimuli were created. The approach was to systematically manipulate scale and orientation characteristics of a background image to create target elements and empirically determine the perceptual salience of these targets with human observers. These data provided baseline performance information. The Neuronal Synchrony Model was then applied to the imagery, another empirical determination of the perceptual salience of the augmented imagery conducted, and a comparison of the effectiveness of the model-augmented imagery with that of the baseline imagery carried out. The creation of the test imagery is described briefly in the next section and in detail in Marchak (2002). The following section provides the methods, results, and analyses of the visual search tasks.

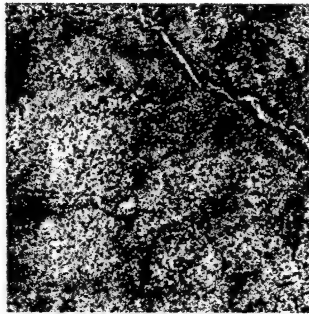
### **TEST IMAGE CREATION**

To test human performance effectiveness of the model, seven different representative background images were used. IKONOS imagery of desert, forest, jungle and mountains as well as SAR imagery of desert, jungle and mountains was selected (Figure 3). These images were chosen to provide a cross section of both environments and sensor types.

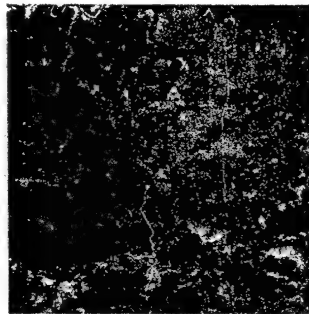
A key factor in selection of the particular Gabor algorithm used in model was the fact that it needed to support both texture analyses for input into the Neuronal Synchrony Model and texture synthesis for the creation of the test images. Test image backgrounds were created by first analyzing the collection of images and determining their Gabor wavelet coefficients. Next, these measurements served as the input to the texture synthesis process of the Gabor algorithm to create synthetic background images based on the



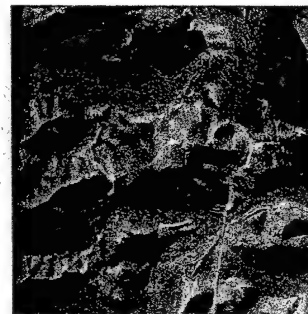
original imagery. As can be seen by comparison of Figures 3 and 4, the synthesized backgrounds captured the fundamental visual characteristics of the original images.



IKONOS Desert – 1m



IKONOS Jungle – 1m



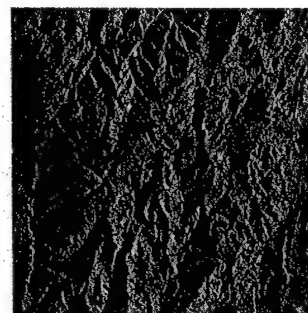
IKONOS Mountains – 1 m



IKONOS Forest – 1m



SAR Desert – 24m

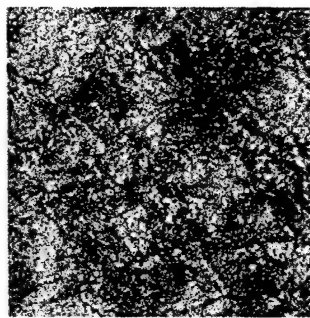


SAR Mountains – 200m

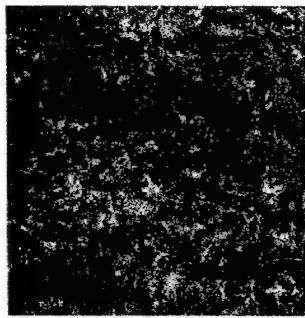


SAR Jungle – 200m

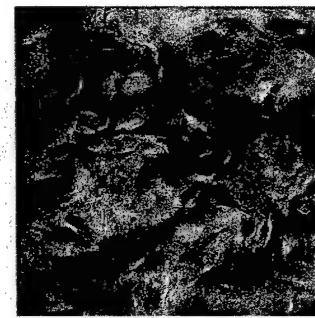
**Figure 3. Original Imagery**



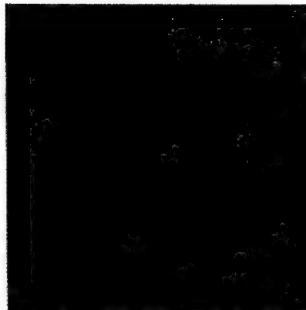
Synthetic IKONOS Desert



Synthetic IKONOS Jungle



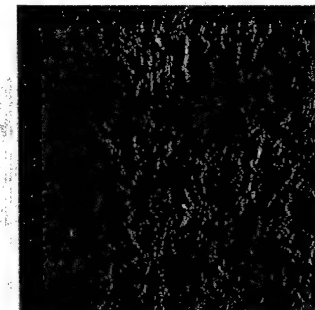
Synthetic IKONOS Mountains



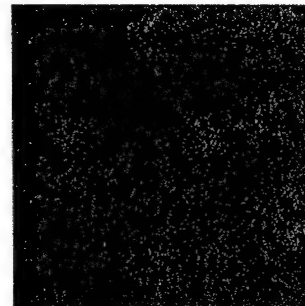
Synthetic IKONOS Forest



Synthetic SAR Desert



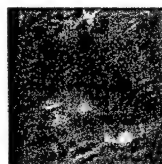
Synthetic SAR Mountains



Synthetic SAR Jungle

**Figure 4. Synthetic Imagery**

The Gabor algorithm was also used to create controlled perturbations of orientation and scale characteristics of the synthetic imagery. Four orientations - 0, 45, 90, and 135 degrees - and four scales - 64, 128, 256, and 512 pixels - were chosen as the basic permutation parameters. To create a permutation, the measured Gabor wavelet coefficients of an image at a particular orientation or scale were multiplied by a constant. The resulting scaled coefficients were then used to synthesize a new image based on that perturbation. Figure 5 shows an example of a synthetic background and the same image with an orientation perturbation at 90 degrees.



Original



90° Orientation Perturbation

**Figure 5. Example of perturbed Imagery**



Based on the results of pilot testing, the values of the perturbations were manipulated to create a final set of 3 levels of salience for each perturbation - low, medium and high - to cover the perceptual span from subtle to obvious. The result was 168 stimuli made up of 3 levels of perceptual salience (low, medium, high), 4 scales (64, 128, 256, 512 pixels), 4 orientations (0, 45, 90, 135 degrees) and 7 backgrounds.

For each background type and perturbation, a test image needed to be created for use in the visual search task. In this approach, a 36-pixel diameter circular pattern was randomly located in a 1000 x 1000 pixel background image, and the corresponding pixels from an identical but perturbed image were placed in the circle (Figure 8). The targets were constrained to be at least 42 pixels from any edge. A complete set of test imagery using all combinations of backgrounds, scales, orientations, and levels of perceptual salience was generated.



**Figure 8. Final test stimulus**

### **PSYCHOPHYSICAL TESTING – SYNTHETIC IMAGERY**

The goal of this task was to conduct a visual search experiment with the stimuli created above in order to determine baseline values of observers' visual search times and eye movement parameters to detect targets embedded in imagery that has not been augmented with the output of the Neuronal Synchrony Model. The resulting data were compared with those obtained using augmented imagery to determine the effectiveness of this approach in aiding image analysis.

### **SUBJECTS**

Twenty subjects, all with normal or corrected-to-normal vision, were paid for their volunteer participation, with ten subjects in each condition, baseline and augmented. The Montana State University Human Subjects Committee approved experimental procedures and protocols.



### APPARATUS

A BARCO Personal Calibrator luminance-calibrated monitor was used to ensure accurate stimulus presentation across subjects and data collection tasks. A Dell Optiplex GX1 running E-Prime, an experimental presentation and data collection software, was used for image presentation. A Dell Optiplex GXi was used to collect data with an ISCAN RK-426 remote eye tracking system.

### PROCEDURES

The E-Prime experimental presentation software was configured into 6 blocks of 35 trials, where each block contained 28 target stimuli and 7 catch trials with no target present. Subjects were first presented with a series of informational displays that described the task and provide examples of hard-to-see and easy-to-see targets, as well as examples of the different background types. This was followed by a series of 35 practice trials to familiarize them with the procedures. A trial began with a fixation point in the center of the screen. When the observer pressed a key, the fixation point was replaced with a test image, the search timer started and eye movement tracking began. The observer had up to 30 seconds to indicate detection or absence of the target by a key press, which recorded the visual search time and stopped eye movement recording. After a short delay the fixation point reappeared and the subject could begin the next trial. The practice trials were identical to the experimental trials with the exception that the data were not analyzed.

Half the subjects performed the task using the original synthetic imagery, and half used the Neuronal Synchrony Model augmented synthetic imagery. Figure 9 shows an example of the original imagery, while Figure 10 shows the same image with the model augmentation. The arrow, not shown in the experiments, indicates the location of the target.

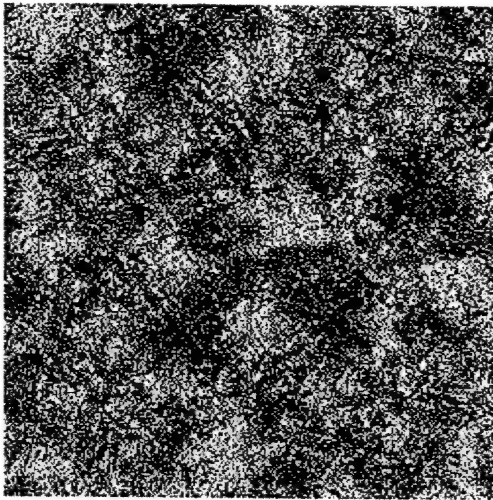


Figure 9. Original synthetic background

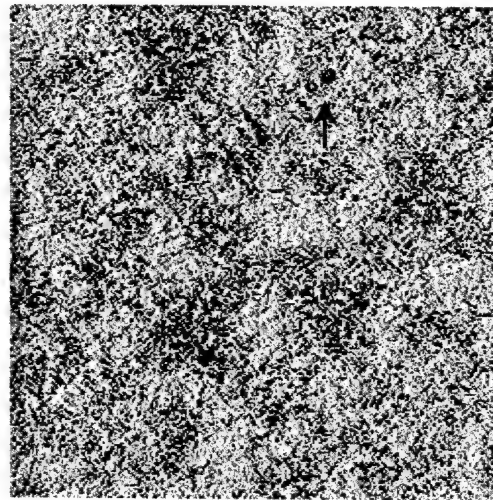


Figure 10. Augmented synthetic background

### RESULTS AND ANALYSES

Accuracy, visual search time, and number of fixations were measured for each of the 3 levels of perceptual salience (low, medium, high), 4 scales (64, 128, 256, 512 pixels), 4 orientations (0, 45, 90, 135 degrees) and 7 backgrounds. Table 1 presents the overall mean accuracy scores across all factors. The first item to note is the overall difficulty of the task; in the Original Imagery condition, subjects averaged about only 43% correct across all background types, while with the augmented condition, accuracy increased to 57%. While





performance in the augmented condition was superior, this difference was not significant ( $\chi^2 = 1.345$ ,  $p = 0.131$ ).

**Table 1. Synthetic Imagery Overall Accuracy**

| Accuracy (Percent Correct) |                   |
|----------------------------|-------------------|
| Original Imagery           | Augmented Imagery |
| 42.6                       | 57.4              |

When the data are examined for interactions among the variables, significant differences emerge. Table 2 shows the accuracy for each background type. There was a significant overall difference in accuracy for background type ( $\chi^2 = 25.85$ ,  $p = .000$ ) and significant interactions of image condition – original versus augmented – and background type. Performance was better with the augmented imagery for the IKONOS Forest ( $\chi^2 = 6.52$ ,  $p = .007$ ), IKONOS Jungle ( $\chi^2 = 44.51$ ,  $p = .00$ ), IKONOS Mountains ( $\chi^2 = 3.49$ ,  $p = .038$ ), and the SAR Jungle ( $\chi^2 = 24.61$ ,  $p = .000$ ).

**Table 2. Synthetic Imagery Accuracy by Background Type**

| Background       | Accuracy (Percent Correct) |                   |
|------------------|----------------------------|-------------------|
|                  | Original Imagery           | Augmented Imagery |
| IKONOS Desert    | 42.1                       | 57.9              |
| IKONOS Forest    | 45.3                       | 54.7              |
| IKONOS Jungle    | 30.1                       | 69.9              |
| IKONOS Mountains | 40.4                       | 59.6              |
| SAR Desert       | 44.1                       | 55.9              |
| SAR Jungle       | 47.6                       | 52.4              |
| SAR Mountains    | 44.0                       | 56.0              |

Table 3 shows the accuracy data for the perturbations in scale versus the perturbations in orientation. There was a significant difference across conditions, with augmentation producing the greatest benefit for orientation perturbations ( $\chi^2 = 15.81$ ,  $p = .000$ ).

Lastly, examining accuracy level as a function of degrees of perceptual salience, Table 4 shows the accuracy data that are significantly different across conditions. The greatest benefit of augmentation is seen for the least perceptually salient stimuli ( $\chi^2 = 9.51$ ,  $p = .001$ ), and the medium perceptually salient stimuli ( $\chi^2 = 3.63$ ,  $p = .034$ ).



**Table 3. Synthetic Imagery Accuracy by Perturbation Type**

| Perturbation | Accuracy (Percent Correct) |                   |
|--------------|----------------------------|-------------------|
|              | Original Imagery           | Augmented Imagery |
| Scale        | 50.7                       | 49.3              |
| Orientation  | 46.2                       | 53.8              |

**Table 4. Synthetic Imagery Accuracy by Perceptual Salience Level**

| Perceptual Salience | Accuracy (Percent Correct) |                   |
|---------------------|----------------------------|-------------------|
|                     | Original Imagery           | Augmented Imagery |
| Low                 | 45.1                       | 54.9              |
| Medium              | 50.8                       | 49.2              |
| High                | 48.7                       | 51.3              |

These accuracy data show that the Neuronal Synchrony Model augmented imagery significantly increase accuracy of performance for particular backgrounds (IKONOS Jungle, IKONOS Mountains, IKONOS Forest, SAR Jungle), differences in orientation, and targets of low and medium perceptual salience.

Turning to the search time and eye movement fixation data, a multivariate analysis of variance (MANOVA) was performed on the correct responses with background type, perturbation type (scale and orientation), and level of perceptual salience as factors. Significant levels of Wilks' Lambda were found between the original and augmented imagery conditions ( $F(2,2709) = 87.446, p = .000$ ) as well as the interaction between background type and condition ( $F(12,5418) = 2.876, p = .001$ ). In addition, level of perceptual salience was also significant ( $F(3,5418) = 3.103, p = .015$ ).

Table 5 shows the means for overall search time and number of fixations by condition. There was a significant difference in search time ( $F(1, 2710) = 151.880, p = .000$ ) and number of fixations ( $F(1, 2710) = 81.954, p = .000$ ), with longer search times and a greater number of fixations for the augmented imagery versus the original imagery.

**Table 5. Synthetic Imagery Mean Overall Search Time and Fixations**

|                     | Original Imagery | Augmented Imagery |
|---------------------|------------------|-------------------|
| Search Time (msec)  | 2578.99          | 4991.95           |
| Number of Fixations | 5.97             | 15.85             |



Table 6 shows mean search time data and Table 7 shows number of fixations as a function of background type for both conditions. There was a significant effect for background type ( $F(6,2710) = 3.059, p = .006$ ) and a significant interaction between background type and condition ( $F(6,1) = 4.549, p = .000$ ) in the search time data. Subjects spent a longer time searching and made more fixations using the augmented imagery than using the original imagery.

**Table 6. Synthetic Imagery Mean Search Time by Background**

| Background       | Search Time (msec) |                   |
|------------------|--------------------|-------------------|
|                  | Original Imagery   | Augmented Imagery |
| IKONOS Desert    | 3170.06            | 4969.01           |
| IKONOS Forest    | 1169.06            | 4948.10           |
| IKONOS Jungle    | 3650.94            | 5148.35           |
| IKONOS Mountains | 3382.11            | 4763.59           |
| SAR Desert       | 2811.56            | 4971.07           |
| SAR Jungle       | 1288.18            | 5136.07           |
| SAR Mountains    | 2581.02            | 5007.45           |

**Table 7. Synthetic Imagery Mean Number of Fixations by Background**

| Background       | Number of Fixations |                   |
|------------------|---------------------|-------------------|
|                  | Original Imagery    | Augmented Imagery |
| IKONOS Desert    | 6.10                | 16.54             |
| IKONOS Forest    | 5.41                | 13.66             |
| IKONOS Jungle    | 5.87                | 17.06             |
| IKONOS Mountains | 6.63                | 15.57             |
| SAR Desert       | 6.85                | 15.18             |
| SAR Jungle       | 4.87                | 16.39             |
| SAR Mountains    | 6.08                | 16.56             |

There were no significant differences between conditions for scale and orientation perturbations. Tables 8 and 9 show mean search time and mean number of fixations as a function of perturbations in scale or orientation, respectively.

**Table 8. Synthetic Imagery Mean Search Time by Perturbation Type**

| Perturbation | Search Time (msecs) |                   |
|--------------|---------------------|-------------------|
|              | Original Imagery    | Augmented Imagery |
| Scale        | 2534.64             | 2623.33           |
| Orientation  | 4987.66             | 4978.64           |

**Table 9. Synthetic Imagery Mean Number of Fixations by Perturbation Type**

| Perturbation | Number of Fixations |                   |
|--------------|---------------------|-------------------|
|              | Original Imagery    | Augmented Imagery |
| Scale        | 5.73                | 6.21              |
| Orientation  | 16.30               | 15.10             |

Tables 10 and 11 show mean search time and mean number of fixations as a function of level of perceptual salience.

**Table 10. Synthetic Imagery Mean Search Time by Perceptual Salience Level**

| Perceptual Salience | Search Time (msec) |                   |
|---------------------|--------------------|-------------------|
|                     | Original Imagery   | Augmented Imagery |
| Low                 | 3199.87            | 5316.55           |
| Medium              | 2457.89            | 4759.31           |
| High                | 2079.21            | 4873.59           |

**Table 11. Synthetic Imagery Mean Number of Fixations by Perceptual Salience Level**

| Perceptual Salience | Number of Fixations |                   |
|---------------------|---------------------|-------------------|
|                     | Original Imagery    | Augmented Imagery |
| Low                 | 5.79                | 15.25             |
| Medium              | 5.93                | 16.14             |
| High                | 6.19                | 15.71             |

As the data above show, subjects spent more time searching and made more fixations to the augmented imagery than to the original imagery. In isolation, this finding suggests that augmentation by the Neuronal



Synchrony Model causes the analyst to search for longer periods of time and make more fixations than searching un-augmented imagery. However, when taken together with the accuracy data, a different interpretation may be reached.

Examination of the overall accuracy data showed that this particular search task was extremely difficult, with subjects reaching an average accuracy of just over 50% across both conditions. In the original imagery conditions, the shorter search times and fewer fixations could be attributed to subjects terminating their search early because the overall homogeneity of the stimuli resulted in false alarms regarding target detection. While there were catch trials with no target present, no mechanism was available, for example, by marking the target location, to verify that when subjects indicated detection that they had actually found the actual target.

In the augmented imagery condition, the highlighting altered the homogeneous nature of the stimuli, thus both serving to better indicate true targets and reduce the potential for falsely identifying targets. Subjects searched longer and made more fixations in the highlighted condition because, at least for certain background types, the highlighting permitted a more effective search and resulted in less frustration than in the original image condition. Informal feedback from the subjects during post-experiment debriefing confirms this explanation, as those in the original imagery condition indicated that they often noticed features of the background that they thought might be targets and that after an initial search, they would indicate a detection if they noticed almost any feature that differed from the background. Subjects in the augmented imagery conditions commented that on some trials the target would "pop-out" of the highlighted background, and on other trials, searching the areas around the highlighting often led to detecting targets that were otherwise difficult to see.

The overall findings of this experiment suggest that imagery augmented by the Neuronal Synchrony Model results in a higher level of detection accuracy and an increase in the effectiveness of the search process by increasing the amount of time spent successfully searching for targets and reducing the tendency to terminate a search early due to lack of feedback. The next section describes refinements made to the model as a result of this experiment as well as a function of applying the model to actual imagery. These refinements are then applied to overhead imagery and the results evaluated with human observers.



### TASK 3 – MODEL REFINEMENT

The Neuronal Synchrony Model developed in Task 1 increased the accuracy and success of visual searches using synthetic imagery. A further test of the model's effectiveness is how well it performs on actual overhead imagery of natural scenes. The goal of this task was to apply the model to representative overhead imagery and refine it to increase its performance in highlighting areas of perceptual salience

#### MODEL APPLICATION TO EXTENDED IMAGERY SET

The model developed in developed in Task 1 was applied to the set of overhead imagery shown in Figure 11. Imagery types included data from aerial imagery, IKONOS, QuickBird, Japanese Earth Resources Satellite (JERS-1), SAR, and SPOT.


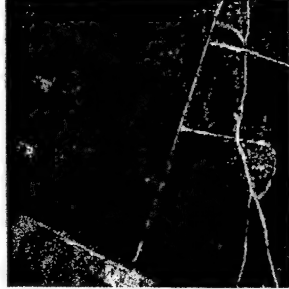
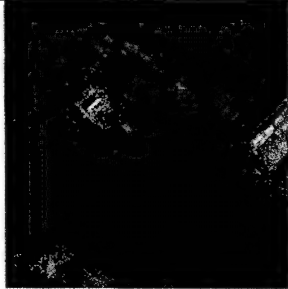



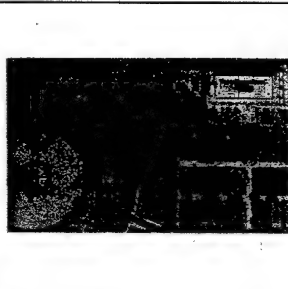
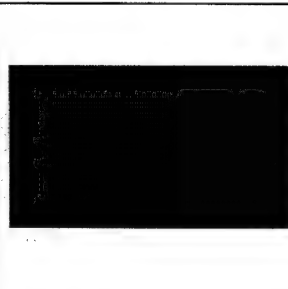
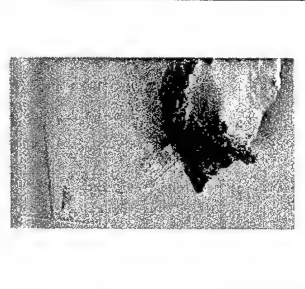
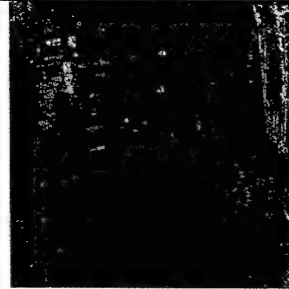
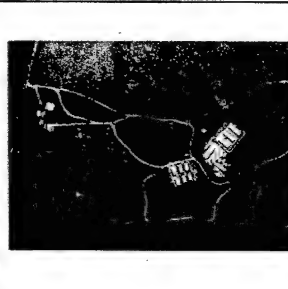
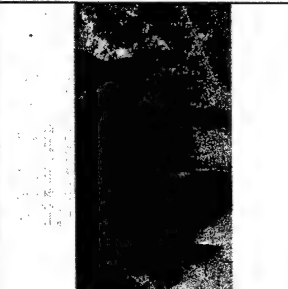


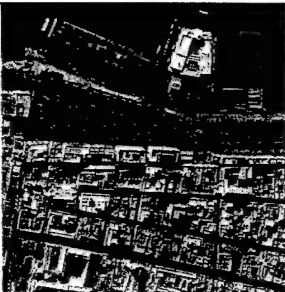





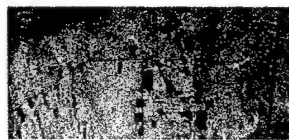
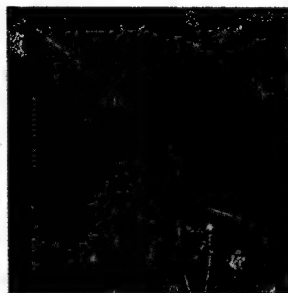
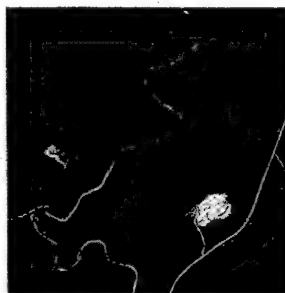
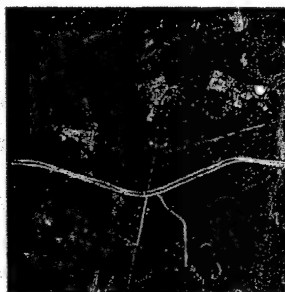
|   |   |  |   |
|---|---|--|---|
|    |    |    |    |
| Aerial 1m Serbia  | Aerial 4m Forest  | IKONOS 1m Iran   | IKONOS 1m Iraq  |
|  |  |  |  |
| IKONOS 1m Kabul   | IKONOS 1m Kandahar  | IKONOS 1m DC   | IKONOS 4m Lingshui  |
|  |  |  |  |
| IKONOS Antarctic  | JERS 15m Japan  | OTH Radar Nicaragua  | LANDSAT 15 Egypt  |

Figure 11. Imagery Set



|   |  |   |  |
|---|--|---|--|
|    |   |   |   |
| QB Kabul 244 cm   | QB Toledo 1m   | QB Trapini  | SAR 1m ABQ   |
|    |   |   |   |
| SAR 1m Hangars  | SAR 1m Isleta  | SAR 1m Licong   | SAR 3m China Lake  |
|  |  |  |  |
| SAR Unknown   | SPOT 10m PA1   | SPOT 10m PA2  | SPOT 10m PA3   |

**Figure 11 (continued). Imagery Set**

Unlike the homogenous, synthetic imagery, actual overhead imagery contains large variations in background structure. While the basic elements of Gabor filtering, facilitation and inhibition calculation functioned well, the particular implementation of neural oscillator was unsuited for this purpose. This is due to the fact that the original LEGION implementation is extremely sensitive to noise in the image because of an oversimplified local coupling scheme (Chen, Wang, & Liu, 2000). To address this shortcoming, an extended LEGION model that contains a weight adaptation scheme and a logarithmic grouping rule in place of the summation and maximization rules used in Terman and Wang (1995) was employed. The next section describes the changes and the implementation of the model.



## UPDATE AND REFINEMENT

The sections below outline the design of the updated and refined Neuronal Synchrony Model and then specify the implementation of the resulting changes.

## MODEL DESIGN

The LEGION model by Wang and Terman (1995) and described in (8) is updated in Chen, et al. ((2000) to:

$$\dot{x}_{ij} = 3x_{ij} - x_{ij}^3 + 2 - y_{ij} + I_{ij}H(p_{ij} - \theta) + S_i + \rho \quad (12a)$$

$$\dot{y}_{ij} = \varepsilon(\gamma(1 + \tanh(x_{ij} / \beta)) - y_{ij}) \quad (12b)$$

where  $H(\bullet)$  is the Heaviside step function, which is defined as  $H(v) = 1$  if  $v \geq 0$  and  $H(v) = 0$  if  $v < 0$ . The Heaviside term provides a mechanism to distinguish between major objects and noisy fragments, with a major object required to contain at least one oscillator located at the center of a large, homogenous region. Noisy fragments do not contain such oscillators. The overall coupling term,  $S_{ij}$  in (12) is then defined as:

$$S_{ij} = S_{ij}^a - W_z H(z - \theta_z) \quad (13)$$

In the previous algorithm (Wang & Terman, 1995), a summation rule is used to summate the weights of neighboring active oscillators and a maximization rule chooses the maximal one. In the revised algorithm (Chen et al., 2000), an alternative grouping rule,  $S_{ij}^a$ , using a logarithmic operation, is defined:

$$S_{ij}^a = \frac{W_{\max} \sum_{(m,n) \in N(i,j;1)} H(x_{mn}) / 1 + |W_{i,j;mn}|}{\log \left( \sum_{(m,n) \in N(i,j;1)} H(x_{mn}) + 1 \right)} \quad (14)$$

where  $x_{mn}$  is the activity of oscillator  $(m,n)$  and  $W_{\max} = I_{\max} - I_{\min}$ , where  $I_{\max}$  and  $I_{\min}$  are the minimal and maximal intensities across the entire image. The remaining terms are as described in the previous section.

To facilitate weight adaptation, fixed connections are assumed in the neural oscillator network. For a given oscillator, the fixed connectivity specifies a group of neighboring oscillators connecting to it. This oscillator corresponds to one pixel in an image, with the lateral attributes associated with it measured from an ensemble of fixed connections. The fixed connection weights are determined directly from the image. For oscillator  $(i,j)$ , the weights of its fixed connection from oscillator  $(k,l)$ ,  $T_{ij,kl}$ , are defined as the difference between the external stimuli received by  $(i,j)$  and  $(k,l)$  in its lateral neighborhood,  $N(i,j,R)$ :





$$T_{ij;kl} = I_{kl} - I_{ij} \quad (15)$$

$I_{ij}$  and  $I_{kl}$  are the intensities of pixel  $(i,j)$  and pixel  $(k,l)$ , respectively, and  $N(i,j,R)$  is defined as

$$N(i,j,R) = \{(k,l) \mid i-R \leq k \leq i+R, j-R \leq l \leq j+R, (k,l) \neq (i,j)\} \quad (16)$$

where  $R$  ( $R \geq 1$ ) is a parameter that determines the size of the lateral neighborhood. For oscillator  $(i,j)$ , the fixed connections exist only in  $N(i,j,R)$  and  $T_{ij;kl} = -T_{ij;kl}$ , where  $(k,l) \in N(i,j,R)$  (Chen et al., 2000).

An oscillator corresponding to a pixel located near the center of a homogenous region tends to have a high potential and thus can lead an oscillator group. For oscillator  $(i,j)$ ,  $\mu_{ij}(1)$ ,  $\sigma_{ij}^2(1)$ ,  $\mu_{ij}(R_p)$ , and  $\sigma_{ij}^2(R_p)$  are calculated using the mean,  $\mu_{ij}$ , and variance,  $\sigma_{ij}^2$ , of the fixed weights, defined as

$$\mu_{ij}(R) = \frac{\sum_{(k,l) \in N(i,j,R)} T_{ij;kl}}{|N(i,j,R)|} \quad (17)$$

$$\sigma_{ij}^2(R) = \frac{\sum_{(k,l) \in N(i,j,R)} (T_{ij;kl} - \mu_{ij}(R))^2}{|N(i,j,R)|} \quad (18)$$

Oscillator  $(i,j)$  is the center of a homogenous region if and only if  $|\mu_{ij}(R_p) - \mu_{ij}(1)| \leq T_\mu$  and  $|\sigma_{ij}^2(R_p) - \sigma_{ij}^2(1)| \leq T_\sigma$ , where  $T_\mu$  and  $T_\sigma$  are two thresholds to reflect homogeneity and  $R_p$  ( $R_p > 1$ ) reflects whether an oscillator is within a homogenous region. A large value of  $R_p$  results in lead oscillators generated from large homogeneous regions.

## MODEL IMPLEMENTATION

An updated LEGION oscillator network was implemented in MATLAB based on the equations in Chen et al. (2002) described above and as implemented in a C program developed by Wang (2001). The program begins by setting up the connection weights, and then selects leader oscillators in homogeneous regions, based on the values of  $T_\mu$ ,  $T_\sigma$ , and  $R_p$ . The results are evaluated to determine if a region exists and the regions are then numbered in a manner similar to that described above.

The Network Filters developed in Task 1 were applied to the imagery in Figure 11. The results were submitted to the revised LEGION algorithm, with the following range of parameter values:  $R_p$  (4-12),  $T_\mu$  (0.1-2),  $T_\sigma$  (1-2),  $R$  (1-2), and  $W_z$  (45-90). Each value was subjectively determined by examining the resultant image output. The output of the LEGION algorithm is a pixel map that assigns a unique number to each region defined in the input image by the correlated oscillations. The resulting region values were each mapped to saturation levels of a yellow hue to produce highlighted regions. All other pixels were assigned to black. The highlights from the oscillator output were then overlaid on the original imagery to indicate perceptually salient regions. Figure 12 shows the resultant augmented images.




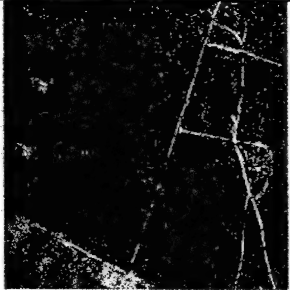
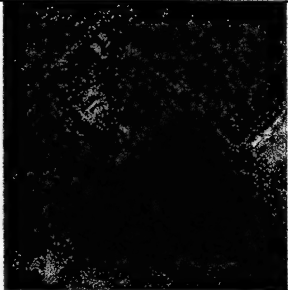


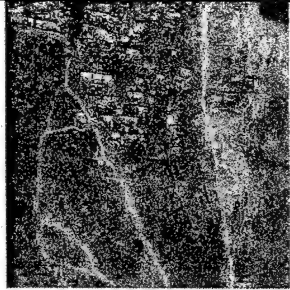
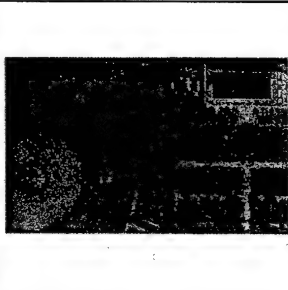
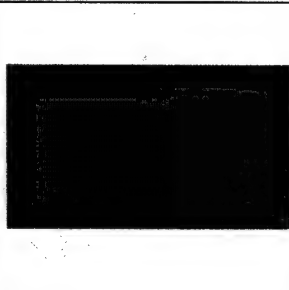
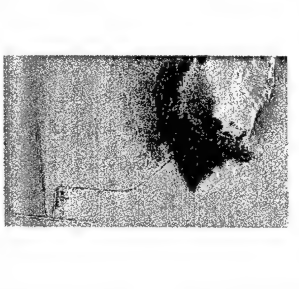
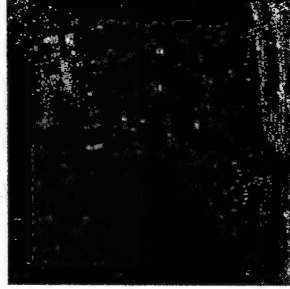
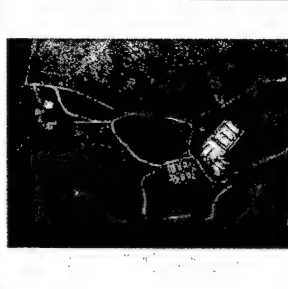
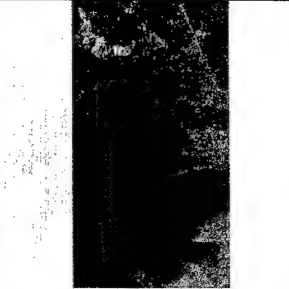



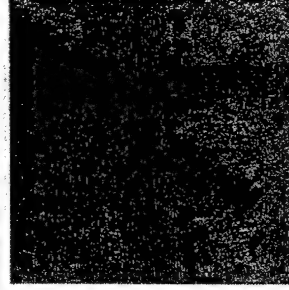
|   |   |  |   |
|---|---|--|---|
|    |    |    |    |
| Aerial 1m Serbia  | Aerial 4m Forest  | IKONOS 1m Iran   | IKONOS 1m Iraq  |
|    |    |    |    |
| IKONOS 1m Kabul   | IKONOS 1m Kandahar  | IKONOS 1m DC   | IKONOS 4m Lingshui  |
|   |   |   |   |
| IKONOS Antarctic  | JERS 15m Japan  | OTH Radar Niceragua  | LANDSAT 15 Eqppt  |
|  |  |  |  |
| QB Kabul 244 cm   | QB Toledo 1m  | QB Trapini   | SAR 1m ABQ  |

Figure 12. Augment Real Imagery




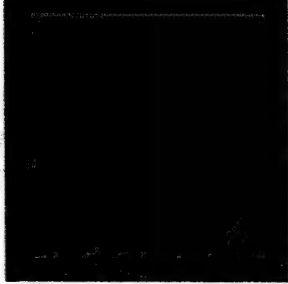

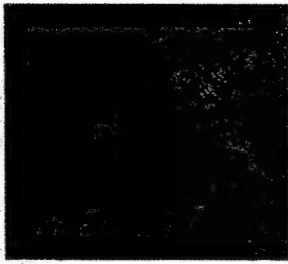
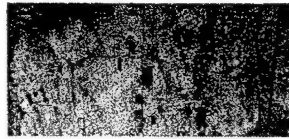
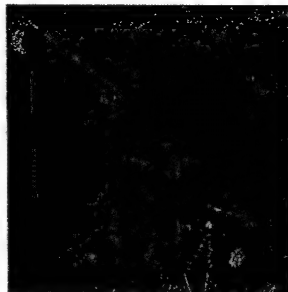

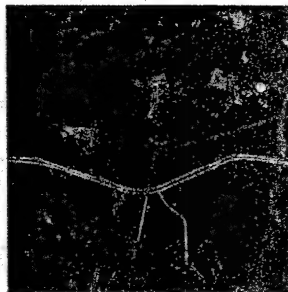
|   |   |  |   |
|---|---|--|---|
|  |  |  |  |
| SAR 1m Hangars  | SAR 1m Isleta   | SAR 1m Licong  | SAR 3m China Lake   |
|  |  |  |  |
| SPOT 10m PA1  | SPOT 10m PA2  | SPOT 10m PA3   | SPOT 10m PA4  |

Figure 12. (continued) Augment Real Imagery

### PSYCHOPHYSICAL TESTING – OVERHEAD IMAGERY

In Task 2, the effectiveness of the Neuronal Synchrony Model was evaluated using synthetic imagery with controlled perturbations. Under this task, a similar methodology was applied to evaluate the effectiveness of the model in aiding human search performance when applied to real overhead imagery. Unlike the first visual search task, it was not possible to define specific targets for detection in the real imagery. In order to produce a quantifiable measure of image contours with some relevance to actual image analysis, imagery was chosen that contained roads and intersections. The subjects' task was to count the number of intersections in an image. If the Neuronal Synchrony Model successfully highlights perceptually salient contours, subjects in the augmented condition should perform better than those in the original imagery condition.

Baseline visual search and eye movement data were collected using the original imagery shown in Figure 11 and compared to performance using the augmented imagery shown in Figure 12. Details of the methods and findings are presented below.

### SUBJECTS

Twenty subjects, all with normal or corrected-to-normal vision, were paid for their volunteer participation, with ten subjects in each condition, baseline and augmented. The Montana State University Human Subjects Committee approved experimental procedures and protocols.



### APPARATUS

A BARCO Personal Calibrator luminance-calibrated monitor was used to ensure accurate stimulus presentation across subjects and data collection tasks. A Dell Optiplex GX1 running E-Prime, an experimental presentation and data collection software, was used for image presentation. A Dell Optiplex GXi was used to collect data with an ISCAN RK426 remote eye tracking system.

### PROCEDURES

The task of the subjects was to count the number of road intersections in each of the images presented. The E-Prime experimental presentation software was configured into 1 block of 29 trials, containing the 24 augmented stimuli and 5 catch trials with no intersections present. Subjects were first presented with a series of informational displays that describe the task and provided examples of different intersection types, as well as examples of the images with roads but no intersections. This was followed by a series of 5 practice trials to familiarize them with the procedures. A trial began with a fixation point in the center of the screen. When the observer pressed a key, the fixation point was replaced with a test image, the search timer started and eye movement tracking began. The observer had unlimited time to count the number of intersections in an image. When the subject had made a determination, they pressed a key. This recorded the visual search time, stopped eye movement recording, and presented a screen that asked "How many intersections were in the image?" Subjects entered their response with a numeric keypad. After a short delay the fixation point reappeared and the subject could begin the next trial. The practice trials were identical to the experimental trials with the exception that the data were not analyzed.

Half the subjects performed the task using the original imagery, and half used the Neuronal Synchrony Model augmented imagery.

### RESULTS AND ANALYSES

Accuracy, visual search time, and number of fixations were measured for each image and image type. Table 12 presents the overall accuracy data across sensor. Similar to the synthetic imagery, the task of accurately counting intersections in real imagery is difficult; in the Original Imagery condition, subjects averaged about 48% correct across all sensor types, while with the augmented condition, accuracy increased to 52%. While performance in the augmented condition was superior, this difference was not significant ( $\chi^2 = 0.62$ ,  $p = .243$ ).

**Table 12. Real Imagery Overall Accuracy**

| Accuracy (Percent Correct) |                   |
|----------------------------|-------------------|
| Original Imagery           | Augmented Imagery |
| 47.8                       | 52.2              |

Table 13 shows the accuracy for each sensor type. The original imagery fared better with radar, LANDSAT, and SPOT imagery, while subjects performed better using the augmented imagery with aerial, IKONOS, and SAR imagery. Only the IKONOS imagery showed a marginally significant difference ( $\chi^2 = 2.88$ ,  $p = .06$ ), favoring the augmented imagery.

**Table 13. Real Imagery Accuracy by Sensor Type**

| Image Type | Accuracy (Percent Correct) |                   |
|------------|----------------------------|-------------------|
|            | Original Imagery           | Augmented Imagery |
| Aerial     | 46.3                       | 53.7              |
| IKONOS     | 44.0                       | 56.0              |
| LANDSAT    | 57.8                       | 41.2              |
| RADAR      | 57.1                       | 42.9              |
| SAR        | 45.0                       | 55.0              |
| SPOT       | 56.3                       | 43.8              |

Turning to the search time and fixation data, Table 14 shows the overall average search time and number of fixations as a function of condition, which are significantly different at the  $p < .10$  level, with  $(F(1,568) = 2.81, p = .094)$  and  $(F(1,568) = 2.83, p = .093)$ , respectively.

**Table 14. Real Imagery Overall Mean Search Time and Fixation Data**

|                     | Original Imagery | Augmented Imagery |
|---------------------|------------------|-------------------|
| Search Time (msec)  | 2075.49          | 1804.91           |
| Number of Fixations | 34.60            | 30.74             |

An examination of search time and number of fixations as a function of sensor and condition, as shown in Tables 15 and 16, shows that mean search time, with exception of SAR imagery, and mean number of fixations were lower for the augmented imagery than the original imagery.

**Table 15. Real Imagery Mean Search Time by Sensor Type**

| Image Type | Search Time (msec) |                   |
|------------|--------------------|-------------------|
|            | Original Imagery   | Augmented Imagery |
| Aerial     | 2463.55            | 1985.26           |
| IKONOS     | 1912.30            | 1667.69           |
| LANDSAT    | 1994.60            | 1645.90           |
| RADAR      | 2152.63            | 1761.63           |
| SAR        | 2036.70            | 2137.03           |
| SPOT       | 1893.17            | 1631.93           |

**Table 16. Real Imagery Number of Fixations by Sensor Type**

| Image Type | Number of Fixations |                   |
|------------|---------------------|-------------------|
|            | Original Imagery    | Augmented Imagery |
| Aerial     | 31.93               | 23.78             |
| IKONOS     | 27.45               | 25.73             |
| LANDSAT    | 42.20               | 36.10             |
| RADAR      | 38.73               | 36.60             |
| SAR        | 34.36               | 31.57             |
| SPOT       | 32.93               | 30.63             |

Overall, the results with real imagery show that while there are no significant differences in accuracy, with the exception of IKONOS imagery, between un-augmented and Neuronal Synchrony Model augmented imagery, both mean search time and number of fixations are lower for the augmented imagery. These finding could be due to several reasons. First, the particular measure of accuracy, as measured by the number of intersections counted, might not be an adequate evaluation of the effectiveness of the augmentation. Since accuracy rates were low in both conditions, the task itself may not have taken advantage of the capabilities of the model. Second, since the model was applied at only one spatial scale and facilitation calculated for only identical orientation filters, the full fidelity of the augmentation may not have been realized. Applying the model across all spatial scales and across all combination of filter and image orientations should result in better performance, as background noise present at only one scale would be reduced. Lastly, better selection of threshold values for the selection of neural oscillators based on objective measures of the characteristics of the imagery could increase identification of additional salient contours.

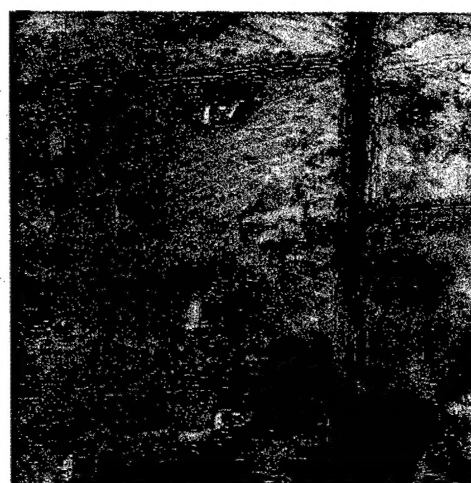
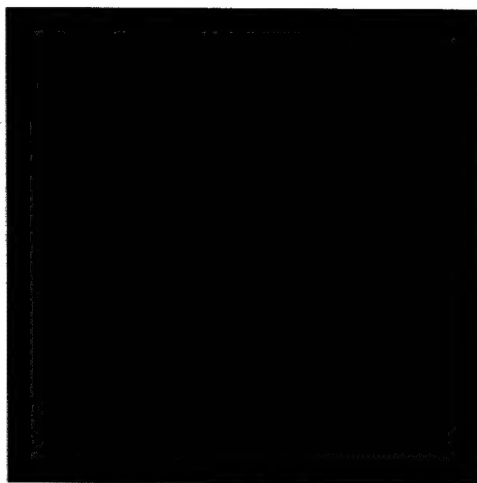


## CONCLUSIONS

The primary objective of this effort was to develop a neurophysiologically based model that characterizes the perceptually relevant contours of interest in overhead imagery of natural scenes in a manner consistent with human performance, utilizing only information available in the image. The goal was to apply the model to highlight contours of potential interest in a manner that augments the analyst's own visual processing. The output of this work was a proof-of-concept demonstration for enhancing the analyst's ability to perform searches on representative imagery.

The results showed that for targets of low to medium perceptual saliency in homogeneous backgrounds, the Neuronal Synchrony Model increase the accuracy of detecting those targets and enhances the ability to search longer with fewer false alarms. When the task involves determining the number of intersections in real overhead imagery, depending on the sensor type, augmentation with the Neuronal Synchrony Model reduces average search time and number of fixations.

In order to demonstrate the feasibility of this approach within the constraints of the current effort, several simplifications were made to the model implementation, include processing at only a subset of the spatial scales and determination of facilitation at only primary orientations. Despite these simplifications, the model nonetheless locates the majority of perceptually relevant contours in an image, and does so in a manner superior to standard edge detection routines. Figure 13 shows the output of the Neuronal Synchrony Model compared to that of a Canny edge detector (Canny, 1986) for an example image.



**Figure 13 (a). Output of Canny edge detector      Figure 13 (b). Output Neuronal Synchrony Model**

As can be seen, the Neuronal Synchrony model ignores much of the distracting background edges, highlighting primarily the perceptible contours of interest.

Overall, the Neuronal Synchrony Models shows great promise in the ability to provide image analysts with a measure of perceptual salience that can guide image search. Further research and expansion of the model will enhance both its performance and utility. The next section outlines topics for further investigation.





## SUGGESTIONS FOR FUTURE RESEARCH

As shown above, the effectiveness of the Neuronal Synchrony Model algorithm proved to vary with the different image types and search tasks. Several factors can be addressed to increase the performance and robustness of the algorithm.

First, as a simplification, the facilitation and inhibition algorithm operated at only one of the four spatial scales of the Gabor filter outputs. Using information from all four spatial scales would result in the ability both to detect relevant contour features across scales as well as to reject background noise that may be present at only one scale. In addition, facilitation was only calculated for orientations of the same value as the facilitation filter. That is, the 45° facilitation filter was only applied to the Gabor output calculated at 45°. Calculating the preferred angle between a particular filter and each orientation, thus summing energy that occurs across all filter outputs, could capture a more accurate level of facilitation.

Second, the revised LEGION method of modeling neuronal oscillation implemented on the real imagery, based on the work of Chen et al., (2000), incorporated only fixed weights to specify connections between oscillators in the network. This fixed connectivity specifies a group of neighboring oscillators connected to an individual oscillator and is determined directly from the image. As described in Chen et al., (2000), dynamic weight connections can also be employed. These dynamic weights capture the instantaneous relationship between two oscillators during weight adaptation, providing noise removal and feature preservation. Further, the particular values for oscillator threshold variables were selected manually and ad-hoc for each individual image. Development of schemes for automatically selecting the currently manually selected parameter values for homogeneity and determination of whether an oscillator is in a homogeneous region would make the process both more objective and more effective.

While the goal of this effort was to detect open contours in imagery, the model on which the algorithm is based (Yen & Finkel, 1998) is especially effective at detecting closed contours, which have a higher degree of perceptual salience. It is possible to extend and expand the current algorithmic implementation of the model to optimize the ability to detect perceptually salient features composed of closed contours, such as buildings and vehicles. This would involve investigating the optimal implementation of the horizontal long-range connections used in the model, as well as tuning the facilitation and inhibition calculations.

Lastly, as discussed above, the method of highlighting used on the imagery was based on applying a single hue and intensity to all background types. A more effective methodology would be to select hue and intensity values based on the contrast and color of the image to be augmented. Investigation of methodologies for adaptively determining the optimal highlighting scheme based on the characteristics of the underlying imagery should increase human performance.



## REFERENCES

- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6), 679-698.
- Chen, K., Wang, D., & Liu, X. (2000). Weight adaptation and oscillatory correlation for image segmentation. *IEEE Transactions on Neural Networks*, 11(5), 1106-1123.
- Deco, G., & Zihl, J. (2000). Neurodynamical mechanism of binding and selective attention for visual search. *Neurocomputing*, 32-33, 693-699.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local Association Field. *Vision Research*, 33, 173-193.
- Gilbert, C. D., Das, A., Ito, M., Kapadia, M., & Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences USA*, 93, 615-622.
- Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: Still alive and well. *Neuron*, 24, 31-47.
- Guindon, B. (1997). Computer-Based Aerial Image Understanding: A Review and Assessment of its Application to Planimetric Information Extraction From Very High Resolution Satellite Images. *Canadian Journal of Remote Sensing*, 23(1).
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's visual cortex. *Journal of Physiology*, 160, 106-154.
- Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological review*, 99(3), 480-517.
- Kisvarday, Z. F., & Eysel, U. T. (1992). Cellular organization of reciprocal patchy networks in layer III of cat visual cortex (area 17). *Neuroscience*, 46, 275-286.
- Kovacs, I., & Julesz, B. (1993). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, 370, 644-646.
- Mallat, S. (1999). *A wavelet tour of signal processing* (2nd ed.). San Diego, CA: Academic Press.
- Marcelja, S. (1980). Mathematical description of the response of simple cortical cells. *Journal of the Optical Society of America*, 70, 1297-1300.
- Marchak, F. M. (2002). *Neurophysiological Based Methods of Guided Image Search: Interim Report* (No. TR-0201-01). Bozeman, MT: Veridical Research and Design.
- Nestares, O., Navarro, R., Portilla, J., & Taberner, A. (1998). Efficient spatial-domain representation of a multi-scale image representation based on Gabor functions. *Journal of Electronic Imaging*, 7, 166-173.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381, 607-609.
- Parent, P., & Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 11, 823-839.
- Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33(993-999).



- Raghu, P. P., & Yegnanarayana. (1997). Multispectral image classification using Gabor filters and stochastic relation neural network. *Neural Networks*, 10(3), 561-572.
- Schmidt, K. E., Goebel, R., Lowel, S., & Singer, W. (1997). The perceptual grouping criterion of collinearity is reflected by anisotropies of connections in the primary visual cortex. *European Journal of Neuroscience*, 9, 1083-1089.
- Singer, W. (1999). Neuronal synchrony: A versatile code for definition of relations? *Neuron*, 24, 49-65.
- Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18, 555-586.
- Sowmya, A., & Trinder, J. (2000). Modelling and representation issues in automated feature extraction from aerial and satellite images. *ISPRS Journal of Photogrammetry & Remote Sensing*, 55, 34-47.
- Sporns, O., Tononi, G., & Edelman, G. M. (1991). Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections. *Proceedings of the National Academy of Sciences USA*, 88, 129-133.
- Terman, D., & Wang, D. (1995). Global competition and local cooperation in a network of neural oscillators. *Physica D*, 81, 148-176.
- von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron*, 24, 95-104.
- von der Malsburg, C., & Buhmann, J. (1992). Sensory segmentation with coupled neural oscillators. *Biological Cybernetics*, 67, 233-242.
- Wang, D. (1994). segment.c. Columbus, OH: DeLiang Wang.
- Wang, D. (2001). cwl.c. Columbus, OH.
- Wang, D., & Terman, D. (1995). Locally excitatory globally inhibitory oscillator networks. *IEEE Transactions on Neural Networks*, 6(1), 283-286.
- Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae*. New York: John Wiley.
- Yen, S.-C., & Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Research*, 38(5), 719-741.